

時系列データの形状認識に基づく言語化への取り組み

土橋 亮子 (指導教員：小林 一郎)

1 はじめに

我々の周囲で観測されるデータの多くは時系列データである。しかし、一般に人は時系列データからではなく、グラフ(視覚)や専門家の解説・文章(言語)を通し、その時系列データの内容を理解していると考えられる。また、時系列データの解釈はその分野の専門家でない限り、迅速かつ確に解釈することは難しい。そのため専門家に代わって、時系列データの内容を解釈し、ユーザに情報を提供をする手法が望まれる。時系列データの振る舞いを解釈する際、人は大局的な動向を視覚で把握し、言葉で理解している。その点に着目し、本研究では株価や為替の動向について時系列データのグラフの形状を認識し、形状にあった適切な言葉を表示できるようなシステムを提案する。

2 日経平均株価テキスト生成システム

2.1 システム概要

先行研究 [1] において開発された「日経平均株価テキスト生成システム」の概要を図 1 に示す。このシステムに円・ドルの為替レートの動向を説明する言語化を行う。

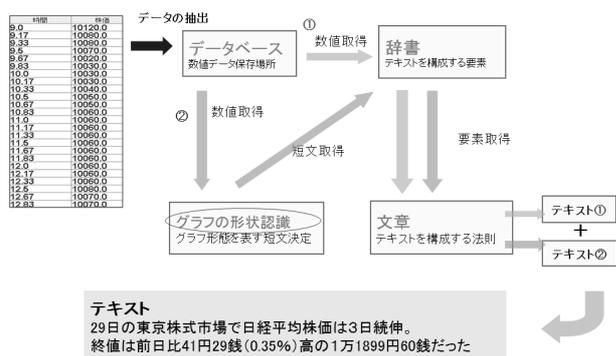


図 1: システムの概要

このシステムによって生成されるテキストは以下の2つのテキストタイプに分類され、タイプごとにテキスト生成の処理の流れが異なる。

テキスト ① : グラフの形状を踏まえることなしに、データベースからの情報のみから生成できるテキスト。

テキスト ② : グラフの形状を踏まえて、かつデータベースからの情報から生成できるテキスト。

本研究においては、テキスト ② の自動生成に着目し、テキスト生成の性能向上およびその評価を行う。以下にシステム各部の説明、および、テキスト ② の生成処理の流れを示す。

2.2 グラフの形状認識

グラフの動向を把握するとき、グラフが「下がって、上がっている」などの形状によって認識される。グラフを視覚的に把握するために、本研究では、線形最小二乗法を用いてグラフの近似曲線を作り、その近似曲線

の振る舞いを捉えることにより、グラフの動向を言語で表す。近似曲線は4次多項式で表現されており、この多項式の次数は、グラフの形状を表現している語彙の実際のコーパス(約1ヶ月分の日経平均株価動向の解説記事)を分析することにより、その最適な次数を4次と導いた。4次多項式が表現する典型的な曲線の全体的な形状を11タイプとし、その形状のパラメータ値のとり方により、さらに13種類の部分形状が導けるとした(図2参照)。

分類	形状	部分形状
type1		
type2		
type3		
type4		

図 2: タイプごとに分類された部分形状(一部)

この分類は、実際のコーパスから抽出されたグラフの挙動を説明するために使われる語彙表現の観点から導いた。任意の全体形状のタイプはどの部分形状を含むかが決まっているため、4次多項式で認識されたグラフの形状は、始めに分類された全体形状の特定のタイプを選別する。次に、その部分形状を数式的に解釈することにより最終的なグラフの形状を認識し、これを説明する適切な言語表現をする(図3参照)。

図3は株価を対象に構築したグラフの形状を言語表現に結びつけられる辞書であり、予備実験により為替のコーパスと出力された言語を比較してみると、為替を対象に言語化する為には先行のままでは不十分であることが判明した。その一つの理由として、為替の変動が円・十銭単位では特に、株価ほど大きな変動示さないことが挙げられる為、微動なグラフの変動に対する言語表現が不足していることが挙げられる。

部分形状	短文+時間帯	特徴
	売りが優勢だった	$ b2-b1 / MAX-MIN >0.4$ $ a1-a2 / max-min <0.7$
	売りが広がった	$ a1-a2 / max-min >0.7$
	売りが優勢になる場面があった	$ b2-b1 / MAX-MIN >0.4$ $ b2-b3 / b2-b1 >0.5$ $ a1-a2 / max-min <0.7$
	中ごろ過ぎにかけて	$ fsh-a2 / max-min <0.7$ $ fsh-a2 / max-min >0.5$
	中ごろに	$ fst-a1 / max-min <0.2$ $ fsh-a2 / max-min <0.2$
	中ごろ過ぎから	$ fsh-a1 / max-min <0.6$ $ fsh-a1 / max-min >0.45$

図 3: 部分形状の数式的解釈とその言語表現

2.3 辞書

そこで、為替特有の語彙辞書を作成し、株価の語彙辞書との比較検討を試みた。辞書は、実際のコーパスとそれに対応する株価動向を示すグラフの部分形状(増加・減少・微動)の対応関係により構築される。

まず為替の動向を示す語彙辞書の作成にあたり、2010年8月2日から2010年9月9日までの114個の円・ドルの為替レートに関する実際のコーパスを分析することにより、グラフの部分形状を適切に表現する文、語彙を収集し、辞書を構築した。

為替動向を表現する本研究で構築した辞書の語彙と、先行研究で作成された日経平均株価動向を表現する辞書の語彙におけるグラフ形状を示す語彙を比較したところ、全く同じ表現の語彙・ほぼ同じ意味の語彙・各々特有の語彙の3つに分類された。

為替動向の時系列データを表現する語彙として、日経平均株価の動向を表現する語彙と類似している語彙も多々あるが、為替動向特有な言葉も多数存在することが分かった。特にグラフの形状が微動の場合を示す語彙表現は、為替に多くみられ追加する必要が大いにあると考えられる(表1,表2参照)。

表 1: 語彙辞書比較内訳

グラフ形状	変更不要	変更必要	株価独自	為替独自
増加	4	9	7	9
減少	2	6	8	10
微動	2	7	4	13
合計	8	22	19	33

表 2: 株価・為替の独特な語彙表現(一例)

形状	株価独自の表現	為替独自の表現
増加	バスケット買いが入った模様 上げ幅は小幅にとどまっている 急速に伸び悩んだ	ドルの買い戻しが散見された 小戻す展開となった 小幅に買われた
減少	売りが膨らんだ 急速に下げ幅を縮小 急速に下げ幅を縮小	大台割れとなり 弱含む場面も見られた 軟化する場面も見られた
微動	買い手控え気分が強まっている 様子見ムードが強まった 底堅さを確認した	膠着状態となっていた 銭程度のレンジ内取引が続いた 売買交錯している

株価動向と為替動向を示す記事の差異はそれぞれの市場の特徴も反映される。株価市場の開場時間は午前9:00~11:00,午後12:30~15:00であるのに対し、東京為替レートは日曜を除き、世界の市場を合すると24時間体制である。コーパス分析より時間帯は、大別すると「東京午前」、「東京午後」、「欧州時間」、「NY時間」の4つに分類されることを確認した。(本研究では東京午前及び東京午後の言語化の取り組みを行う。)

構築された辞書は、上述のような為替レート説明に特有の表現、および、図3に示すようにグラフの形状を数式的に解釈したものが為替独特の語彙や文も含め対応するようにシステム内に実装されている。

辞書内には部分形状で表現できる短文が約70種類(例:「ドル買いでじりじりと値を上げた」「下げ幅拡大した」「膠着状態となっていた」)、時間帯が5種類(例:「東京午前」「序盤」)、接続詞が4種類(例:「その後」「しかし」)が登録されている。

2.4 文法

テキストは、短文、時間帯、接続詞の適切な組み合わせ規則により生成される。その例を以下に示す。

- 時間帯によって先頭に「東京午前は」、「東京午後は」をつける。
- 部分形状によっては、時間帯によって「序盤」、「現

在は」、などが短文の前につけられる。

2.5 実行例

図4では、「2010年9月6日」と入力すると、以下のようなテキストが生成された。

タイプ①テキスト

「6日のドル円為替レートは反落。終値は前週末比0円20銭(0.24%)安の8432円00銭だった。184円近くまで上昇する場面があった。」

タイプ②テキスト

「東京午前は、売りが広がった。下げ渋った。東京午後には、膠着した取引が続いた。」

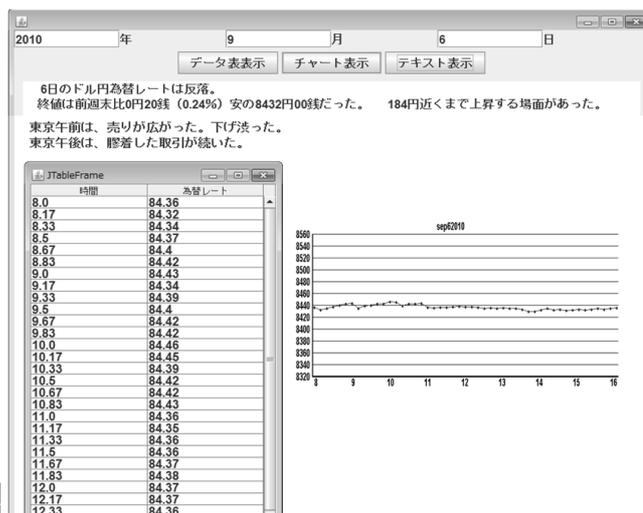


図 4: システムの実行例

3 おわりに

本研究において、先行研究[1]で構築された株価動向を言語化するシステムを基に、言語化の対象となる時間帯の変更等、為替レート(円ドル)の動向を示す言語化を可能にする機能の追加を進めた。

実際に為替レートの値を先行システムに代入し、集めたコーパスと比較検証をしてみると、為替レートの動向を示す言語表現は、株価動向を示す言語表現と類似しているものもあるが、株価にはない、為替独自の表現、動向が存在することが分かった。そのため、為替に適した語彙辞書を作成し、株価の語彙と比較を行うことで、為替と株価の違いを発見し、為替の言語表現にあったシステムを構築することができた。

今後の課題として、為替の言語表現をより精密にできるようにする、円台で推移している等、グラフ形状から生成された文章に、具体的な数値を入れより自然な言語表現にする、そして株価と為替の変動に関する相互関係を言語化するなど、複数の時系列データの相関関係などを言語するつもりである。

参考文献

- [1] 小林一郎, 渡邊千明, 奥村奈穂子: グラフとテキストの協調による知的な情報提示手法 日経平均株価テキストとグラフの提示を例にして, 情報処理学会論文誌, 48(3), pp.1058-1070, 2007
- [2] 加藤, 松下: 動向情報の要約・可視化から情報編集へ, 第21回人工知能学会全国大会, 2H5-11, (2007).